

Selenge Tulga

selengetu@gmail.com | [Portfolio](#) | [Linkedin](#) | [Github](#)

Education

Master of Data Science - University of Rochester

December 2024

Bachelor of Computer Science - Mongolian University of Science and Technology

June 2017

Professional experience

Data Science Intern - UPMC Motion Lab (Rochester, NY)

May 2024 -Sep 2024

- Automated data pipeline with Python scripting, saving the team over 10 hours of manual work per week
- Built and maintained scalable data pipelines using Python to process and integrate 100,000+ patient records for business-critical analytics and reporting, ensuring data accuracy and regulatory compliance

Database Engineer - Zevtronic LLC

Jan 2020 -July 2023

- Designed and implemented ETL processes and batch pipelines to transform and integrate large datasets from multiple healthcare sources, improving data processing speed by 70%.
- Developed ETL processes using SQL and Python, ensuring data quality and accessibility for analytics. Leveraged AWS Athena for query optimization and data integration across multiple data sources.

Senior Software engineer - Ulaanbaatar Railway JVC

Jan 2022-July 2023

- Led the development of data pipelines to process and transform data from 60+ systems, supporting various business functions including operations, logistics, and finance. Improved data retrieval speed by 200% through optimized data architecture.
- Architected scalable and high-performance data storage solutions, improving retrieval speeds by 200%

Software engineer - Ulaanbaatar Railway JVC

July 2017-Dec 2021

- Managed and maintained data integrity from over 40 locomotive signal systems, employing rigorous testing protocols to uphold system reliability and data quality.
- Developed robust backend systems for locomotive repair and passenger booking applications,, which resulted in a 300% revenue increase through improved operational efficiency and user experience.

Independent Projects

Tweet Sentiment Analysis using Databricks Spark

- Transformed 200K tweets using PySpark and Delta Lake architecture on Databricks for real-time analysis.
- Designed a real-time sentiment analysis pipeline with Spark Streaming and Medallion architecture.
- Deployed a pre-trained HuggingFace model via MLflow for scalable real-time sentiment classification.

Exploring sides of Two conflicts On Reddit and Weibo

- Extracted and analyzed data from user comments on Reddit and Weibo using Python and various APIs
- Applied Roberta and Llama models for sentiment analysis and war conflict studies.
- Utilized regression models, fine-tuned Llama 2, and visualized data with Tableau and PowerBI.

Certification

[Databricks Certified Data Engineer Associate](#)

[AWS Certified Data Engineer Associate](#)

[Databricks Certified Data Engineer Professional](#)

Skills

Programming Languages: Python, Java, SQL, PHP, R

Databases: MySQL, MSSQL, Oracle, PostgreSQL, MongoDB, Redshift, DynamoDB

Cloud Computing: AWS (EC2, S3, Lambda, DynamoDB, Athena, Presto), Azure, Google Cloud

Data Engineering Tools: Apache Spark, Airflow, Kafka,, AWS(Glue), Google Cloud,

Other: Data Governance, Data Quality, ETL/ELT, Linux, REST APIs, Git, Llama